



Identifying optimal survey-based algorithms to distinguish diabetes type among adults with diabetes

Jennifer G. Nooney^{a,*}, M. Sue Kirkman^b, Kai McKeever Bullard^c, Zachary White^a, Kristi Meadows^a, Joanne R. Campione^a, Russ Mardon^a, Gonzalo Rivero^a, Stephen R. Benoit^c, Emily Pfaff^d, Deborah Rolka^c, Sharon Saydah^c

^a Westat, Rockville, MD, USA

^b Division of Endocrinology and Metabolism, Department of Medicine, University of North Carolina, Chapel Hill, NC, USA

^c Division of Diabetes Translation, Centers for Disease Control and Prevention, Atlanta, GA, USA

^d NC TraCS Institute, University of North Carolina, Chapel Hill, NC, USA

ARTICLE INFO

Keywords:

Algorithms
Type 1 diabetes
Diabetes surveillance
Surveillance methodology

ABSTRACT

Objectives: Surveys for U.S. diabetes surveillance do not reliably distinguish between type 1 and type 2 diabetes, potentially obscuring trends in type 1 among adults. To validate survey-based algorithms for distinguishing diabetes type, we linked survey data collected from adult patients with diabetes to a gold standard diabetes type. **Research design and methods:** We collected data through a telephone survey of 771 adults with diabetes receiving care in a large healthcare system in North Carolina. We tested 34 survey classification algorithms utilizing information on respondents' report of physician-diagnosed diabetes type, age at onset, diabetes drug use, and body mass index. Algorithms were evaluated by calculating type 1 and type 2 sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) relative to a gold standard diagnosis of diabetes type determined through analysis of EHR data and endocrinologist review of selected cases.

Results: Algorithms based on self-reported type outperformed those based solely on other data elements. The top-performing algorithm classified as type 1 all respondents who reported type 1 and were prescribed insulin, as "other diabetes type" all respondents who reported "other," and as type 2 the remaining respondents (type 1 sensitivity 91.6%, type 1 specificity 98.9%, type 1 PPV 82.5%, type 1 NPV 99.5%). This algorithm performed well in most demographic subpopulations.

Conclusions: The major federal health surveys should consider including self-reported diabetes type if they do not already, as the gains in the accuracy of typing are substantial compared to classifications based on other data elements. This study provides much-needed guidance on the accuracy of survey-based diabetes typing algorithms.

Introduction

Type 1 and type 2 diabetes are distinct clinical conditions with different average ages of onset, management strategies, associated complications, and patient outcomes, but with overlap in many phenotypic factors. An estimated 90–95 percent of all diabetes cases in adults are type 2 diabetes with most of the remainder being type 1 [1]. The comparative rarity of type 1 means that surveillance of diabetes is largely driven by trends in type 2 diabetes. It is therefore important for surveillance systems to be able to distinguish diabetes type to support type-specific analyses of morbidity, mortality, medical care costs, and health-related quality of life. Currently, the large federal surveys that

the nation relies upon to monitor trends in diabetes prevalence and incidence are unable to reliably distinguish between types. The purpose of this study is to identify the most accurate combination of survey questions for identifying diabetes type.

The National Health and Nutrition Examination Survey (NHANES), which allows for diabetes and prediabetes prevalence estimation, including undiagnosed cases [2,3], and the Behavioral Risk Factor Surveillance System (BRFSS), which allows state-level diabetes prevalence estimation for the Centers for Disease Control and Prevention's (CDC) Diabetes Atlas [4], do not include self-reported diabetes type. Currently, efforts to use these surveys to distinguish type rely upon items assessing diabetes drugs used, body mass index (BMI), and age at

* Corresponding author at: Westat, 1600 Research Blvd., Rockville, MD 20850, USA.

E-mail address: jennifernooney@westat.com (J.G. Nooney).

<https://doi.org/10.1016/j.jcte.2020.100231>

Received 4 February 2020; Received in revised form 26 May 2020; Accepted 18 June 2020

2214-6237/ © 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

diagnosis. To our knowledge, there are no studies assessing the validity of these approaches using a gold standard diagnosis type for survey respondents.

The California Health Interview Survey (CHIS) as well as the Survey on Living with Chronic Disease in Canada (SLCDC) include a question on self-reported diabetes type. CDC and the National Institute of Diabetes and Digestive and Kidney Disease also supported the inclusion of questions related to diabetes type and insulin requirements in a supplement to the 2016 and 2017 National Health Interview Survey (NHIS). However, it is unknown how accurate self-reported diabetes type is in these populations, and it is unknown whether typing algorithms including self-reported type are more accurate than those using other information to distinguish type.

Although several studies have employed items available from national surveys to develop algorithms for classifying diabetes type, none of these studies have validated their questions and algorithms against the contents of the medical record as the gold standard [5–11]. This is a critical step for confirming the utility of self-reported diabetes type and the performance of survey algorithms based on other questions relevant to determining type. In this paper, we report the results of a validation study of survey-based algorithms for identifying diabetes type using data collected from adult patients with diabetes from a large healthcare system compared to a gold standard diabetes type derived from collection of structured and unstructured data from patients' electronic health records (EHRs).

Methods

Sample selection and gold standard diabetes type classification

As shown in Fig. 1, we selected a sample of 2,500 adult patients from the UNC Health Care System (UNCHCS) who were highly likely to have diabetes based on three years of EHR data (10/1/2014 – 9/30/2017). We required at least one visit to a primary care or endocrinology clinic within the prior 18 months. We modified the straw man algorithm used by Klompas and colleagues [12] to find likely diabetes cases, incorporating diagnosis codes, laboratory tests, and diabetes drugs. Appendix Table A1 provides the specifications of our straw man algorithm. To facilitate detection of the rarer type 1 diabetes, we over-sampled probable type 1 diabetes by identifying patients with two or more type 1 diagnosis codes on separate occasions OR one type 1 code

on the patient's problem list AND no outpatient prescription for non-insulin hypoglycemic medications. The sample was further stratified by age, sex, and race/ethnicity to facilitate testing of algorithms in demographic subpopulations. Our target of 2,500 cases was designed to yield 1,000 completed surveys, assuming a 40% response rate (achieved by the similarly designed North Carolina BRFSS) [13].

We developed a gold standard diabetes type for each case through analysis of both structured and unstructured EHR data. Unstructured EHR data from chart notes were abstracted by trained staff, including a nurse practitioner, endocrinologist, and research assistants, to identify age at diagnosis, historical use of insulin and oral antidiabetic medications, and other elements not available in diagnosis codes and other structured data points. To produce a preliminary gold standard diagnosis, we applied two quantitative models independently to each case – a decision tree and a weighting equation. The decision tree used sequential rules to classify patients based on clinical factors (see Appendix Fig. A1). Straightforward cases were classified at the top of the tree based on type-congruent diabetes diagnosis codes (the same codes used in our straw man algorithm) and type-consistent medication use in the patients' medical records. More complex cases were classified using additional data elements (e.g., age at diagnosis, type-specific laboratory tests, family history) moving down the tree. In contrast, the weighting equation simultaneously considered twelve clinical factors using a scoring system in which clinical characteristics weighed towards or against Type 1 or Type 2 (see Appendix Table A2). Both methods permitted a classification of "indeterminate." Any cases for which the two methods did not agree, plus any cases scored as indeterminate by one or both methods, were reviewed by an endocrinologist to make a final determination of type (N = 282).

A total of 2,465 cases were determined to have diabetes, were not deceased, and received a gold standard diabetes type at the end of the process.

Survey design and operations

We designed a telephone survey incorporating widely used and validated survey items from national and international surveys including NHIS, BRFSS, and NHANES. Small adaptations were made to adjust items for telephone administration; question wording and response options for key items in the survey are shown in Appendix Table A3. We made one modification to the wording used in the 2016–2018

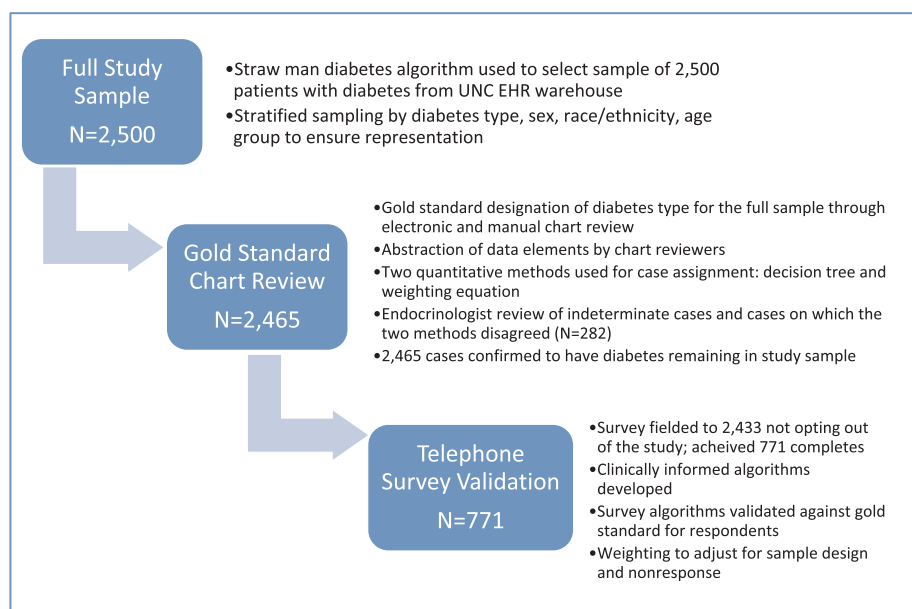


Fig. 1. Study design.

NHIS survey for self-reported diabetes type. NHIS asked “What type of diabetes do you have?” with response options of “type 1, type 2, and other.” We asked “According to your doctor or other health professional, what type of diabetes do you have? Is it type 1, type 2, or some other type? If you don’t remember or weren’t told, that’s OK.” This revised question wording was included in the 2019 NHIS. In addition to self-reported diabetes type, we asked whether respondents were on insulin or non-insulin diabetes medications, the timing of insulin use (whether within 1 year of diagnosis, whether ever stopped taking insulin), age at diabetes diagnosis, weight and height (for body mass index), and whether respondents had ever experienced diabetic ketoacidosis.

Eight patients completed cognitive interviews by phone to test the survey. Once finalized, the survey was translated into Spanish to minimize language barriers for Spanish-speaking respondents. The survey was programmed as a Computer Assisted Telephone Interview (CATI). We sent a pre-notification letter with an opt-out option to the sample of 2,465 and fielded the survey to those who did not call in to opt out of the study (32 people opted out; final N = 2,433). Trained CATI interviewers administered the survey between November 2018 and February 2019. Respondents completing the interview were provided a gratuity of \$25.00. The survey achieved a response rate of 32.7%, collecting data from 771 diabetes patients.

Weighting and analytic design

Base weights were developed using the demographic and type 1 oversample classifications of all UNCHS diabetes patients meeting the straw man criteria and eligible for sampling (N = 41,614). The base weight for each case was calculated as the inverse of its selection probability, which varied by race/ethnicity, gender, age, and type 1 oversample eligibility. The base weight adjusted the sample to reflect the distribution of cases eligible for sampling, since we enriched our sample with non-whites, younger individuals, and presumptive type 1 cases. Importantly, the weights correct for a very high oversampling rate for presumptive type 1 cases, which made up 26.9% of the study sample but only 5 percent of all cases meeting straw man criteria. The oversampling optimized our ability to design algorithms for the rarer type 1 diabetes while the weighting allowed us to test our algorithms in the “real world” where diabetes cases are dominated by the more common type 2 diabetes. Base weights were further adjusted for survey non-response patterns and calibrated through raking to our known sample totals by race/ethnicity, gender, age, and presumptive type 1. The raking ensured that weighted totals of respondents who completed the survey would match the totals of the sampling frame for all of the sampling strata.

Our analytic strategy involved linking survey data with the gold standard diabetes type for each case and implementing case definitions and diabetes typing algorithms from the literature as well as extensions of these algorithms. To identify the best performing algorithms, we computed weighted validity statistics (sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV)) for type 1 and type 2 diabetes. Several aggregate validity statistics were also examined, including the proportion correctly classified by the algorithm and an average of type 1 sensitivity and PPV. All data management and validity testing was done using SAS version 9.4 (SAS Institute Inc., Cary, NC).

Algorithms were selected for testing if they appeared in the literature and, for newly created algorithms, if they answered clinically-informed questions about the utility of additional survey items for identifying type 1 diabetes. Examples include testing the added value in requiring that insulin use begin within a year of diabetes diagnosis, whether insulin should be required to be continuous, and whether relaxing the continuous insulin requirement to include potential type 1 cases who stopped insulin only during the one-year “honeymoon” period after diagnosis is advantageous. The 33 resulting algorithms

clustered into three groups based on the algorithms’ most important, or anchoring, data element: self-reported diabetes type, medication use, or age at first diagnosis.

We also utilized a machine learning modeling approach, which resulted in one additional algorithm, to ensure we had not missed an important algorithm for distinguishing type that had not been tested in prior studies. We built a conditional inference tree that optimized sequential partitioning (splitting) of selected survey variables, creating a decision tree [14]. This modeling technique is advantageous because it produces a result that can be easily implemented by users of the major federal health surveys. The stopping criterion for the tree is whether a potential additional split passes a statistical significance test. A further split is not made if all potential splits are not statistically significant. We relied on repeated cross-validation using the same data set to choose the optimal level of complexity of the model. In each repetition, we randomly split the data into 5 folds. We sequentially used 4 of them to fit a model, and we measured the performance of each combination on the remaining fold. The full process of 5-fold cross-validation was repeated 5 times. The average performance across all 25 model fits was used to select the optimal algorithm. Once fitted, the resulting decision tree was coded for the full sample, weighted, and evaluated with the validity statistics discussed above. The conditional inference tree was built in R version 3.6.0 [15].

Algorithm performance statistics

We used several general principles to guide the assessment of the algorithms and the selection of the optimal algorithm. First, because the prevalence of type 1 diabetes is so much lower, it is especially important to correctly identify these cases. Although each validity statistic provides information on an algorithm’s performance in a different area, we prioritized type 1 sensitivity, type 1 PPV, and type 2 specificity when selecting the top performing algorithm. We also examined the estimated weighted prevalence of each type in comparison to the weighted prevalence of the gold standard as a check on the algorithms’ face validity. Finally, we report a weighted percentage of cases that were correctly classified by the model. This measure incorporates, and weighs heavily toward, the performance of algorithms in identifying type 2. Ideally, the top-performing algorithm excels in correctly classifying all cases of diabetes.

To apply these principles, each of the algorithms was evaluated by examining its type 1 sensitivity and type 2 specificity. Any algorithms with values below 70% on either of these metrics were eliminated from further consideration, as were algorithms with type 1 or type 2 PPV less than 30%. Because PPV is highly sensitive to prevalence, when weights are applied the type 1 PPV for many algorithms is low. We therefore chose a less restrictive cut off for PPV as compared with sensitivity and specificity.

A complication for clinicians and researchers is the inability to classify some cases as type 1 or type 2. Among the survey sample, the gold standard classification included only 16 cases categorized with indeterminate and “other” types of diabetes. Therefore, the survey algorithms were limited in the capacity to classify those with indeterminate diabetes. Although some of the survey-based algorithms also generated classifications of “other,” these other and indeterminate cases often operated as a source of error, which reflects a naturally-occurring difficulty in estimating diabetes prevalence by type.

The Institutional Review Boards of both Westat and the UNC approved the study design and protocol, and the research team ensured full compliance with all applicable restrictions on the handling and transfer of Protected Health Information and Personally Identifiable Information.

Results

Characteristics of the sample reporting a diabetes diagnosis and

Table 1
Characteristics of survey respondents reporting diabetes (N = 698).

Characteristics	Unweighted Sample %	Weighted Sample %
Gold Standard Diabetes Type		
Type 1	32.2	5.3
Type 2	65.5	94.3
Other/indeterminate	2.3	0.4
Diabetes Drugs in EHR		
% using insulin	66.5	42.1
% using non-insulin	54.4	80.2
% on no diabetes drugs	4.6	6.7
Other EHR Information		
Mean onset age, years	33.3	46.8
Mean count of type 1 Dx codes	4.6	0.6
Mean count of type 2 Dx codes	9.0	9.6
Age, years		
18–44	29.1	7.6
45–64	42.6	44.8
65+	28.4	47.6
Sex		
Female	51.7	53.9
Male	48.3	46.1
Race/ethnicity		
Non-Hispanic white	32.7	65.6
Non-Hispanic black	28.2	26.1
Hispanic	23.1	3.3
Non-Hispanic other	16.0	5.0
Education		
Less than high school	17.2	8.6
High school graduate	24.9	28.0
Some college or more	58.0	63.4

Notes: Gold Standard Diabetes Type refers to the “true” diabetes type determined through the use of structured and unstructured electronic health records data, as opposed to the type reported by respondents on the survey. EHR = electronic health record; Dx = diagnosis.

completing subsequent items on diabetes type are shown in Table 1. Almost one-third (32.2%) of respondents had a gold standard diagnosis of type 1 diabetes. When weights are applied to adjust results to the UNCHS system’s diabetes population, only 5.3% of the sample is gold standard type 1. In keeping with this adjustment for our oversampling of type 1 diabetes patients, when compared with unweighted statistics, weighted statistics show a smaller percentage using insulin, a larger percentage using non-insulin diabetes medication, a higher mean age of onset, and fewer type 1 diagnosis codes in their medical records. Table 1 also illustrates our oversampling of non-whites and younger individuals. About 9.5 percent of respondents reported they did not have diabetes and did not complete diabetes-specific survey items necessary for determining type. Nearly all of these respondents (97%) were type 2 by the gold standard. Characteristics of those not reporting diabetes are shown in Appendix Table A4.

Appendix Table A5 contains the full set of 34 algorithms tested in this study, organized by the three anchoring data elements of self-reported type, medication use, or diagnosis age. It also includes our rationale for ruling algorithms out. Diagnosis age-based algorithms tended to perform poorly when the requirement was that diagnosis occur before age 30, with type 1 sensitivities less than 80%. While lifting the age to 40 increased type 1 sensitivity to 84.2%, the positive predictive values for type 1 and type 2 were less than 30%. Although algorithms requiring insulin with no non-insulin diabetes medications slightly outperformed those based on current insulin with or without additional diabetes medications, we note that use of non-insulin medications for glycemic control among patients with type 1 diabetes is currently under review by the FDA and the subject of numerous randomized controlled trials [16]. While insulin-only algorithms may identify patients with type 1 diabetes well today, in the next few years they may become less reliable for diabetes typing analyses.

The final candidate algorithms were based on self-report (Table 2) and medication use (Table 3). The model-based conditional inference

tree is included with self-report algorithms. Despite access to a range of information about sample members, including their use of medication, BMI, age at diagnosis, family history, the presence of autoimmune diseases, and episodes of diabetic ketoacidosis (DKA) and severe hypoglycemia, the model optimized with the use of one variable: self-reported diabetes type. The model selected the split as follows: respondents who report type 1 are classified as type 1, and all others are classified as type 2.

Algorithms including self-reported diabetes type outperformed algorithms based on medication use by a wide margin, when evaluated based on the average of type 1 sensitivity and type 1 PPV. We found that adding the restriction that self-report of type 1 must also be accompanied by current insulin use improved our metrics slightly. In particular, type 1 PPV increased substantially, while type 1 specificity improved modestly, as did most of the type 2 statistics. The final three columns in Table 2 evaluate whether additional restrictions, such as starting insulin within a year of diagnosis and using insulin continuously, improve algorithm performance. These restrictions reduced sensitivity and PPV for type 1, while improving type 1 specificity very slightly and degrading many of the type 2 statistics.

The top-performing candidate algorithm classifies as type 1 anyone who self-reports type 1 and is also on insulin, classifies as “other” anyone who reports another form of diabetes, and classifies as type 2 the remainder of the sample. It requires only three survey items: 1) self-report of diabetes, 2) self-report of diabetes type, and 3) current use of insulin. The algorithm’s type 1 prevalence closely matched the gold standard’s type 1 prevalence (5.9% and 5.3%, respectively) and exceeded all other algorithms in the average of type 1 sensitivity and PPV (87.1%; type 1 sensitivity 91.6%, type 1 specificity 98.9%, type 1 PPV 82.5%, type 1 NPV 99.5%).

The top-performing algorithm also performed well across most study subgroups (Table 4 and Appendix Table A4). The average of type 1 sensitivity and PPV was above 80% in most groups. Black and Hispanic respondents, as well as those without a college education, had average of type 1 sensitivity and PPV below 80%. The algorithm was particularly good at classifying type 1 among women, those aged 18–44 and 45–64, and whites, based on the average of type 1 sensitivity and PPV. We found that the overall top-performing algorithm was also the top-performing algorithm in most subgroups. For blacks, the conditional inference tree-based algorithm achieved a slightly higher average type 1 sensitivity and PPV (66.8%). For respondents with less than a high school education, the top performing algorithm assigned type 1 if the respondent was currently on insulin, started insulin within a year of diagnosis, and was diagnosed younger than age 40 (type 2 otherwise; type 1 sensitivity and PPV average of 55.2%).

Conclusions

Distinguishing diabetes type in national and state-based surveys is important for public health prevention and management strategies. To date, researchers have used survey items on self-reported diabetes type, age at onset, and diabetes drug use to distinguish type without guidance on how accurate these approaches are. The present study provides this much-needed guidance. To our knowledge, it is the first large-scale effort to compare survey responses with a known diabetes type to provide guidance to national surveillance efforts on: 1) items to include in the major federal health surveys to distinguish type, and 2) how to combine items to assign the most accurate diabetes type.

We evaluated 34 algorithms for distinguishing type against a gold standard derived from respondents’ medical records and found that an algorithm based on self-reported type and insulin use performed the best. The algorithm was highly sensitive and specific to both type 1 and type 2 diabetes. We tested algorithms based on anchoring data elements of self-reported type, diabetes drug use, and age at onset. Algorithms based on self-reported type performed better than algorithms based on other anchoring data elements in our study.

Table 2
Prevalence and Performance of Self Report-Based Algorithms (N = 698).

Conditional Inference Tree: Self-report of type 1 = type 1; all others are type 2		Self-report of type 1 + current insulin use = type 1; self-report of Other is Other; all others are type 2	Self-report of type 1, current insulin use, started ins in 1 year of Dx without stopping = type 1; self-report of Other is Other; all others are type 2	Self-report of type 1, current insulin use, ins in 1 year of Dx without stopping except in 1st year = type 1; self-report of Other is Other; all others are type 2
Type 1 Sensitivity (%95 CI)	92.7% (88.6%–96.9%)	91.6% (87.4%–95.8%)	81.6% (75.2%–88.0%)	79.0% (72.4%–85.6%)
Type 1 PPV (%95 CI)	72.9% (56.2%–89.5%)	82.5% (70.8%–94.3%)	83.4% (70.7%–96.0%)	82.9% (70.0%–95.9%)
Type 1 Specificity (%95 CI)	98.1% (96.5%–99.7%)	98.9% (98.0%–99.8%)	99.1% (98.3%–99.9%)	99.1% (98.3%–99.9%)
Type 1 NPV (%95 CI)	99.6% (99.3%–99.8%)	99.5% (99.3%–99.8%)	99.0% (98.6%–99.4%)	98.8% (98.4%–99.3%)
Average of type 1 sensitivity and PPV	82.8%	87.1%	82.5%	81.0%
Type 1 Prevalence: Algorithm	6.7%	5.9%	5.2%	5.0%
Type 1 Prevalence: Gold Standard	5.3%	5.3%	5.3%	5.3%
Type 2 Sensitivity (%95 CI)	98.2% (96.6%–99.8%)	99.0% (98.2%–99.9%)	99.2% (98.4%–100.0%)	99.2% (98.4%–100.0%)
Type 2 PPV (%95 CI)	99.3% (99.0%–99.6%)	99.4% (99.1%–99.7%)	98.9% (98.4%–99.3%)	98.7% (98.3%–99.1%)
Type 2 Specificity (%95 CI)	88.6% (83.8%–93.4%)	90.1% (85.9%–94.4%)	80.8% (74.6%–87.0%)	78.4% (72.0%–84.8%)
Type 2 NPV (%95 CI)	74.6% (57.7%–91.6%)	84.9% (73.3%–96.6%)	86.1% (73.6%–98.5%)	85.7% (72.9%–98.4%)
% Correctly classified	97.5%	98.3%	97.9%	97.8%

Notes: Type 1 Prevalence: Algorithm prevalence refers to the prevalence of type 1 generated by applying the algorithm to the sample. Type 1 Prevalence: Gold Standard refers to the prevalence of type 1 according to the “true” diabetes type determined through the use of structured and unstructured electronic health records data. The first column describes the algorithm that resulted from a conditional inference tree modeling technique; the tree optimized with the use of a single variable, self-reported diabetes type. All estimates are weighted to the full diabetes population of UNC Healthcare. Bold numbers signify the primary metric used to evaluate algorithms.

PPV = positive predictive value; NPV = negative predictive value, Dx = diagnosis; CI = confidence interval.

Table 3
Prevalence and Performance of Drug-Based Algorithms in 2019, UNC Health Care System (N = 698).

	Current insulin use within 1 yr of Dx = type 1; all others are type 2	Current insulin use within 1 yr of Dx without stopping = type 1; all others are type 2	Current insulin use within 1 yr of Dx without stopping except in first year = type 1; all others are type 2	Current insulin use, started insulin within 1 year of Dx, Dx age less than 40 = type 1; all others are type 2
Type 1 Sensitivity (%95 CI)	84.8% (78.7%–91.0%)	82.2% (75.9%–88.6%)	83.8% (77.6%–90.0%)	75.8% (69.0%–82.5%)
Type 1 PPV (%95 CI)	32.9% (22.0%–43.8%)	42.2% (29.6%–54.8%)	42.0% (29.7%–54.2%)	61.9% (46.1%–77.6%)
Type 1 Specificity (%95 CI)	90.4% (85.7%–95.0%)	93.7% (90.5%–96.9%)	93.5% (90.4%–96.7%)	97.4% (95.7%–99.1%)
Type 1 NPV (%95 CI)	99.1% (98.7%–99.5%)	99.0% (98.5%–99.4%)	99.0% (98.6%–99.5%)	98.6% (98.2%–99.1%)
Average of type 1 sensitivity and PPV	58.9%	62.2%	62.9%	68.8%
Type 1 Prevalence Algorithm	13.6%	10.3%	10.6%	6.5%
Type 1 Prevalence Gold Standard	5.3%	5.3%	5.3%	5.3%
Type 2 Sensitivity (%95 CI)	90.6% (85.9%–95.2%)	93.9% (90.7%–97.1%)	93.7% (90.5%–96.9%)	97.6% (95.9%–99.3%)
Type 2 PPV (%95 CI)	98.9% (98.5%–99.4%)	98.7% (98.3%–99.2%)	98.8% (98.4%–99.3%)	98.5% (98.0%–98.9%)
Type 2 Specificity (%95 CI)	83.3% (77.2%–89.4%)	79.9% (73.5%–86.4%)	81.7% (75.4%–87.9%)	74.7% (68.0%–81.3%)
Type 2 NPV (%95 CI)	34.7% (23.2%–46.1%)	44.0% (30.9%–57.0%)	43.8% (31.1%–56.6%)	65.4% (48.8%–81.9%)
% Correctly classified	89.9%	92.9%	92.8%	96.1%

Notes: Type 1 Prevalence: Algorithm prevalence refers to the prevalence of type 1 generated by applying the algorithm to the sample. Type 1 Prevalence: Gold Standard refers to the prevalence of type 1 according to the “true” diabetes type determined through the use of structured and unstructured electronic health records data. All estimates are weighted to the full diabetes population of UNC Healthcare. Bold numbers signify the primary metric used to evaluate algorithms.
PPV = positive predictive value; NPV = negative predictive value; Dx = diagnosis; CI = confidence interval.

Table 4
Type 1 Prevalence and Performance of Top-Performing Survey Algorithm in Subpopulations (N = 698).

	Type 1 Sensitivity (95% CI)	Type 1 PPV (95% CI)	Type 1 Specificity (95% CI)	Type 1 NPV (95% CI)	Average of Type 1 Sensitivity and PPV	Type 1 Prevalence – Algorithm	Type 1 Prevalence – Standard
Female (N = 361)	92.9% (87.4%–98.4%)	95.0% (89.9%–100.0%)	99.7% (99.4%–100.0%)	99.6% (99.3%–99.9%)	93.9%	5.0%	5.1%
Male (N = 337)	90.2% (83.7%–96.8%)	72.1% (53.2%–90.9%)	98.0% (96.1%–99.8%)	99.4% (99.0%–99.8%)	81.1%	6.9%	5.5%
Age 18–44 (N = 203)	88.6% (80.5%–96.7%)	90.2% (80.8%–99.5%)	96.9% (93.6%–100.0%)	96.4% (93.3%–99.4%)	89.4%	23.9%	24.3%
Age 45–64 (N = 297)	94.9% (89.2%–100.0%)	82.8% (62.1%–100.0%)	99.0% (97.5%–100.0%)	99.7% (99.4%–100.0%)	88.9%	5.8%	5.0%
Age 65+ (N = 198)	89.9% (81.2%–98.6%)	72.6% (47.6%–97.6%)	99.1% (98.1%–100.0%)	99.7% (99.5%–99.9%)	81.3%	3.1%	2.5%
Non-Hispanic White (N = 228)	92.0% (86.8%–97.2%)	93.6% (86.0%–97.2%)	99.6% (99.0%–100.0%)	99.5% (99.1%–99.8%)	92.8%	6.1%	6.2%
Non-Hispanic Black (N = 197)	84.2% (74.2%–94.2%)	39.1% (13.6%–64.6%)	97.2% (94.2%–100.0%)	99.7% (99.4%–99.9%)	61.7%	4.6%	2.1%
Hispanic (N = 161)	82.7% (60.5%–100.0%)	57.4% (16.4%–98.5%)	98.8% (96.9%–100.0%)	99.7% (99.2%–100.0%)	70.1%	2.8%	2.0%
Less than HS (N = 119)	72.5% (38.3%–100.0%)	12.7% (0.0%–33.9%)	94.1% (86.4%–100.0%)	99.7% (99.3%–100.0%)	42.6%	6.7%	1.2%
High School (N = 172)	87.7% (78.3%–97.2%)	68.5% (41.8%–95.2%)	98.4% (96.6%–100.0%)	99.5% (99.1%–99.9%)	78.1%	4.8%	3.7%
Some college or more (N = 401)	93.0% (88.2%–97.9%)	97.4% (94.1%–100.0%)	99.8% (99.6%–100.0%)	99.5% (99.2%–99.9%)	95.2%	6.3%	6.6%

Notes: Type 1 Prevalence: Algorithm prevalence refers to the prevalence of type 1 generated by applying the algorithm to the sample. Type 1 Prevalence: Gold Standard refers to the prevalence of type 1 according to the “true” diabetes type determined through the use of structured and unstructured electronic health records data. All estimates are weighted to the full diabetes population of UNC Healthcare. Bold numbers signify the primary metric used to evaluate algorithms.
PPV = positive predictive value; NPV = negative predictive value; CI = confidence interval.

Our results suggest that the major federal health surveys may want to consider including self-reported diabetes type in future iterations of the surveys, if they do not already. The cost of adding this question may be worth the gain in accuracy in identifying diabetes type. Self-reported diabetes type is a powerful marker for the gold standard type, indicating that our survey respondents were generally aware of their condition and able to report it accurately. Bullard and colleagues [11] noted that about 5% of NHIS diabetes respondents reported type 1 but also did not report being on insulin. In our sample, only 14 respondents (2%) reported type 1 without also reporting current insulin use. The lower proportion with incongruous responses may be a result of the rewording of the question on self-reported type, or better patient-provider communication on management of type 1 diabetes within our sample.

While statistical testing from the conditional inference tree did not show great improvement from adding insulin use to self-report for our population, the inclusion of insulin use may be important when moving to a community-based survey population that includes individuals who do not receive regular medical services and who may not have been told or do not remember their diabetes type. We required at least one visit to a primary care or endocrinology clinic within the prior 18 months to ensure that sufficient data were available in the EHR to make a gold standard classification for each case. Beyond this requirement, our sample was extremely diverse, incorporating a majority of non-white respondents and representation across all age groups.

The inclusion of restrictions on timing of insulin onset relative to diagnosis, and continuous use of insulin (less any “honeymoon” lapses in the first year of diagnosis), did not substantially improve the performance of our recommended algorithm. Although they increased the algorithm’s type 1 specificity, they reduced the measure’s type 1 sensitivity and PPV as well as the weighted percent correctly classified. This may be partially explained by adult onset type 1 diabetes, which is often indolent in its presentation and patients may go longer than a year before needing insulin. Though there may be applications for algorithms that emphasize type 1 specificity, for general surveillance by type these items may not be necessary.

Our top-performing algorithm was also the top performer in most study subgroups, but the conditional inference tree-based algorithm (which uses self-reported type alone) outperformed it slightly among blacks. Interestingly, the top-performing algorithm among those with less than a high school education did not use self-reported diabetes type at all and instead relied on more objective elements including insulin use and diagnosis age. Those with lower levels of education may not understand their diabetes as well, and as a result their self-report of type may be less accurate. Indeed, the survey included a question on confidence in self-report of diabetes type, and we found much lower levels of confidence among those with less education. These results suggest that researchers focused on diabetes among those with less education may benefit from using an alternative algorithm to distinguish type.

The major limitations of our study include a sample of patients from a single health system in one state and a final response rate lower than the rate targeted to power subpopulation analysis. The number of responses also impacted our ability to classify indeterminate or other cases of diabetes. Although the top-performing algorithm does permit an outcome of “Other” diabetes type, because only 16 survey respondents had a gold standard determination of indeterminate/other, we were not able to adequately model this outcome. Small sample sizes limited our ability to gauge algorithm performance for some targeted subgroups that were rare in the UNCHS system’s diabetes patient population. In particular, the limited pool of type 1 candidates among blacks and Hispanics hampered our ability to judge algorithm performance in these groups.

While the study team curated the sample carefully during the gold standard development to ensure that all sample members had diabetes, the survey did not lead with this assumption and offered respondents

the opportunity to state that they did not have diabetes (and therefore to skip all portions of the survey specific to diabetes). The 9.5% of respondents who reported they did not have diabetes agrees with the work of prior studies finding moderate agreement between self-report of disease and the medical record [17,18] and cautioning that self-reported rates of many chronic diseases are lower than those obtained through clinical data [19]. We excluded these cases when testing the typing algorithms, but we note that these cases are by definition part of the typing error in the major federal surveys.

This study was the first to comprehensively evaluate diabetes typing algorithms from survey data against a gold standard derived from patients’ medical records. These findings help validate the accuracy of survey questions and their combinations to differentiate type 1 from type 2 diabetes.

Author contributions

J.G.N. collected and analyzed data, contributed to discussion, and wrote the manuscript. M.S.K., E.P., G.R., and J.R.C. analyzed data, contributed to discussion, and reviewed/edited the manuscript. K.M. and Z.W. collected data, contributed to discussion, and reviewed/edited the manuscript. K.M.B., S.R.B., S.S., D.R., and R.M. contributed to discussion and reviewed/edited the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by Grant Number DP006327-01, funded by the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention or the Department of Health and Human Services.

J.G.N. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

The authors thank Derek Scott, Howard King, and a group of dedicated CATI interviewers for their assistance in collecting the survey data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jcte.2020.100231>.

References

- [1] Genuth SM, Palmer JP, Nathan DM. Classification and diagnosis of diabetes. In: Cowie CC, Casagrande SS, Menke A, Cissell MA, et al. (Eds.), Chapter 1 in *Diabetes in America*, 3rd ed., Bethesda, MD, National Institutes of Health, NIH Pub No. 17-1468, 2018.
- [2] Zhang X, Geiss LS, Cheng YJ, Beckles GL, Gregg EW, Kahn HS. The missed patient with diabetes: how access to health care affects the detection of diabetes. *Diab Care* 2008;31:1748–53.
- [3] Zhang X, Bullard KM, Gregg EW, Beckles GL, Williams DE, Barker LE, et al. Access to health care and control of ABCs of diabetes. *Diab Care* 2012;35:1566–71.
- [4] Centers for Disease Control and Prevention. Interactive Diabetes Surveillance System website <https://gis.cdc.gov/grasp/diabetes/DiabetesAtlas.html>. Accessed September 15, 2019.
- [5] Maddigan SL, Feeny DH, Majumdar SR, Farris KB, Johnson JA. Understanding the determinants of health for people with type 2 diabetes. *Am J Public Health* 2006;96(9):1649–55.
- [6] Ng E, Dasgupta K, Johnson JA. An algorithm to differentiate diabetic respondents in the Canadian Community Health Survey. *Health Rep* 2008;19(1):71.
- [7] Menke A, Orchard TJ, Imperatore G, Bullard KM, Mayer-Davis E, Cowie CC. The prevalence of Type 1 Diabetes in the United States. *Epidemiology*

- 2013;24(5):773–4.
- [8] Ng E, Vanderloo SE, Geiss L, Johnson JA. Concordance between self-report and a survey-based algorithm for classification of type 1 and type 2 diabetes using the 2011 population-based survey on living with chronic diseases in Canada (SLCDC)-diabetes component. *Can J Diabetes* 2013;37(4):249–53.
 - [9] Lessem SE. Distinguishing diabetes type [dissertation]. Madison: University of Wisconsin, Madison; 2015.
 - [10] Xu G, Liu B, Sun Y, Du Y, Snetselaar L, Hu F, et al. Prevalence of diagnosed type 1 and type 2 diabetes among US adults in 2016 and 2017: population based study. *BMJ* 2018;362:k1497.
 - [11] Bullard KM, Cowie CC, Lessem SE, et al. Prevalence of diagnosed diabetes in adults by diabetes type – United States, 2016. *MMWR* 2018;67(12):359–61.
 - [12] Klompas M, Eggleston E, McVetta J, Lazarus R, Li L, Platt R. Automated detection and classification of type 1 versus type 2 diabetes using electronic health record data. *Diab Care* 2013;36:914–21.
 - [13] North Carolina Center for Health Statistics. Technical Notes 2017: Behavioral Risk Factor Surveillance System website <https://schs.dph.ncdhhs.gov/data/brfss/2017/technical.htm> Accessed September 30, 2019.
 - [14] Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: a conditional inference framework. *J Comput Graph Statist* 2006;15(3):651–74.
 - [15] R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria 2014. <http://www.R-project.org/> Accessed September 30, 2019.
 - [16] Danne T, Garg S, Peters A, Buse J, Mathieu C, Pettus J, et al. International consensus on risk management of diabetic ketoacidosis in patients with type 1 diabetes treated with Sodium-Glucose Cotransporter (SGLT) Inhibitors. *Diab Care* 2019;42(6):1147–54.
 - [17] Okura Y, Urban LH, Mahoney DW, Jacobsen SJ, Rodeheffer RJ. Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. *J Clin Epidemiol* 2004;57(10):1096–103.
 - [18] Martin LM, Leff M, Calonge N, Garrett C, Nelson DE. Validation of self reported chronic conditions and health services in a managed care population. *Am J Prev Med* 2000;18(3):215–8.
 - [19] Schenker N, Raghunathan TE, Bondarenko I. Improving on analyses of self-reported data in a large-scale health survey by using information from an examination-based survey. *Stat Med* 2010;29(5):533–45.